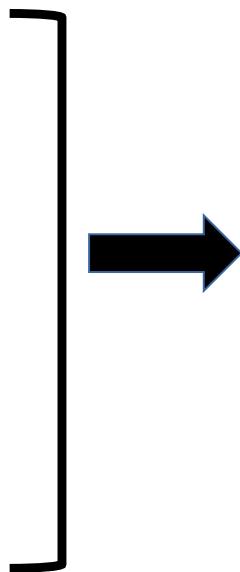


Phylogenomic Analyses of Host Species

- Evolutionary history of species
- Population structure and phylogeography
- Historic and ongoing patterns of migration and gene flow
- Genetic diversity within and between genomes
- Detection of Inbreeding
- Detection of hybridization
- Endangered or relic species, subspecies and populations



Applications

- Identification of genes involved in disease resistance and progression
- Definition of patterns of evolution and selection within targeted genes
- Identification of variants linked with disease pathogenicity in host
- Predictive effects of pathogen emergence in naïve host populations or species

The issue for phylogenomics: making sense of millions of DNA variants within and between species

Table 1 | Genetic variation summary by species and subspecies

Genus	Scientific name species/ subspecies	Common name	N	Mean coverage	Fixed sites to human reference	No. of SNVs*	Mean SNVs per individual*	No. of singletons†	Ancestry informative markers (AIMs)‡	$N_e (10^{-3})\$$
<i>Homo</i>	<i>Homo sapiens</i>	Non-African	6	18.3	386,974	5,887,443	2,639,546	1,379,448	12,316	9.7–19.5
		African	3	20.9	632,253	6,309,453	3,203,178	2,448,454	12,316	13.9–27.9
		Humans	9	19.2	224,660	9,172,573	3,061,604	3,827,902	NA	13.1–16.2
<i>Pan</i>	<i>Pan troglodytes ellioti</i>	Nigeria–Cameroon	10	16.7	25,017,403	12,605,585	4,816,435	2,695,109	2,213	18.5–37.0
	<i>Pan troglodytes schweinfurthii</i>	Eastern	6	28.7	25,126,506	11,264,879	4,843,530	2,228,396	1,265	19.7–39.5
	<i>Pan troglodytes troglodytes</i>	Central	4	23.8	25,080,750	11,820,858	4,983,933	3,948,347	619	24.4–48.7
	<i>Pan troglodytes verus</i>	WesternII	4	27.3	26,832,247	4,729,933	2,411,501	1,481,079	145,548	9.8–19.5
	<i>Pan troglodytes</i>	Common Chimpanzees	24	22.5	24,087,088	27,153,659	5,693,903	10,352,931	149,645	30.9–61.8
<i>Gorilla</i>	<i>Pan paniscus</i>	Bonobos	13	27.5	27,068,299	8,950,002	2,738,755	3,159,889	NA	11.9–23.8
	<i>Gorilla beringei graueri</i>	Eastern lowland	3	22.8	34,537,496	3,866,117	2,578,328	484,482	317,028	12.2–24.3
	<i>Gorilla gorilla diehli</i>	Cross river	1	17.6	35,553,861	2,585,360	2,585,360	165,482	35,693	14.9–29.8
	<i>Gorilla gorilla gorilla</i>	Western lowlandII	23	17.8	31,602,620	17,314,403	6,410,662	2,797,388	19,902	26.8–53.5
<i>Pongo</i>	<i>Pongo abelii</i>	Gorillas	27	18.3	31,376,203	19,177,989	6,492,831	3,447,352	372,623	28.4–56.9
	<i>Pongo pygmaeus</i>	Sumatran	5	28.7	62,880,923	14,543,573	7,263,256	5,681,303	1,132,808	27.5–55.0
		Bornean	5	25.8	64,249,235	10,321,213	5,763,354	3,555,596	1,132,808	19.5–39.0
		Orangutans	10	27.3	60,661,869	24,309,920	9,338,148	6,409,648	NA	42.3–84.6
		All	83	23.0	83,954,672	88,764,143	NA	NA	NA	NA

Speciation, taxonomy, systematics

Avian phylogeny

What makes a species?

Biological Species Concept:

The biological species concept defines a species as members of populations that actually or potentially interbreed in nature, not according to similarity of appearance. Although appearance is helpful in identifying species, it does not define species.

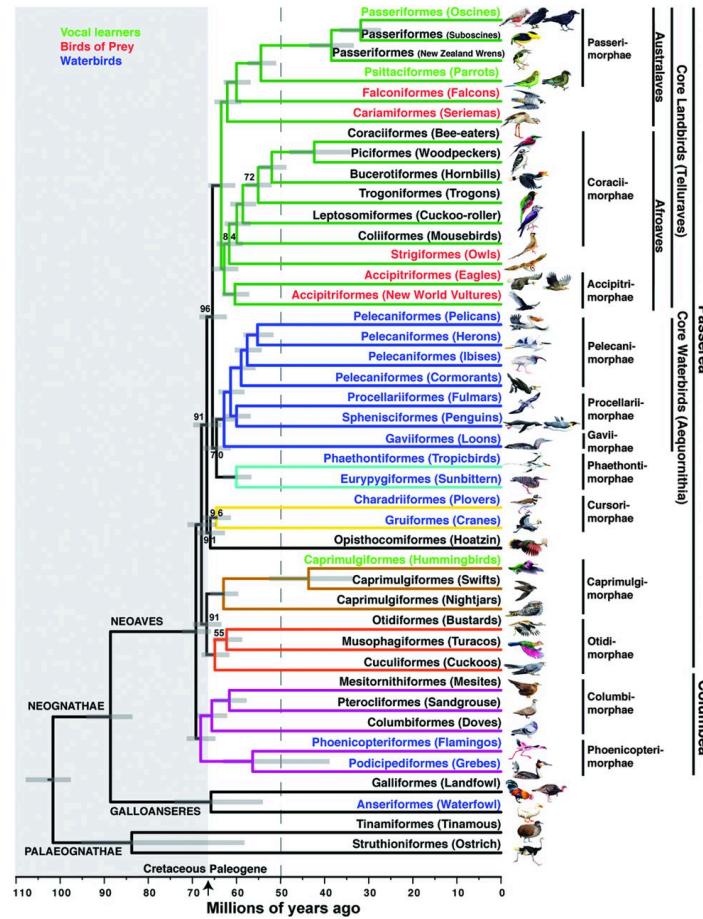
Can Whole Genomes Define a Species?

Vertebrate Genomes ~30,000 genes on average

Evaluate gene trees to support or define expected patterns of speciation

Multi-Species Coalescence

Whole genome data are analyzed under the multispecies coalescent (MSC) that incorporates the genome-wide heterogeneity of gene trees to accurately infer speciation history.



Erich D. Jarvis et al. Science 2014;346:1320-1331

Factors that can lead to inaccurate species trees....

- Recombination
- Gene conversion
- Gene duplication
- Gene loss
- Selection
- Insufficient elapsed time since last shared common ancestor
- Ancestral polymorphisms
- Incomplete lineage sorting (discordant gene trees)
- Saturation of sites (multiple hits)
- Long branch attraction (inadequate taxon sampling)
- Inadequate models of amino acid or nucleotide substitution

...now hold new insights about species divergence made possible through whole genome data

Example Baleen Whale Phylogenomics:

Whole genome sequences

Genomes of rorquals are characterized by contradicting genealogies for their central divergence.

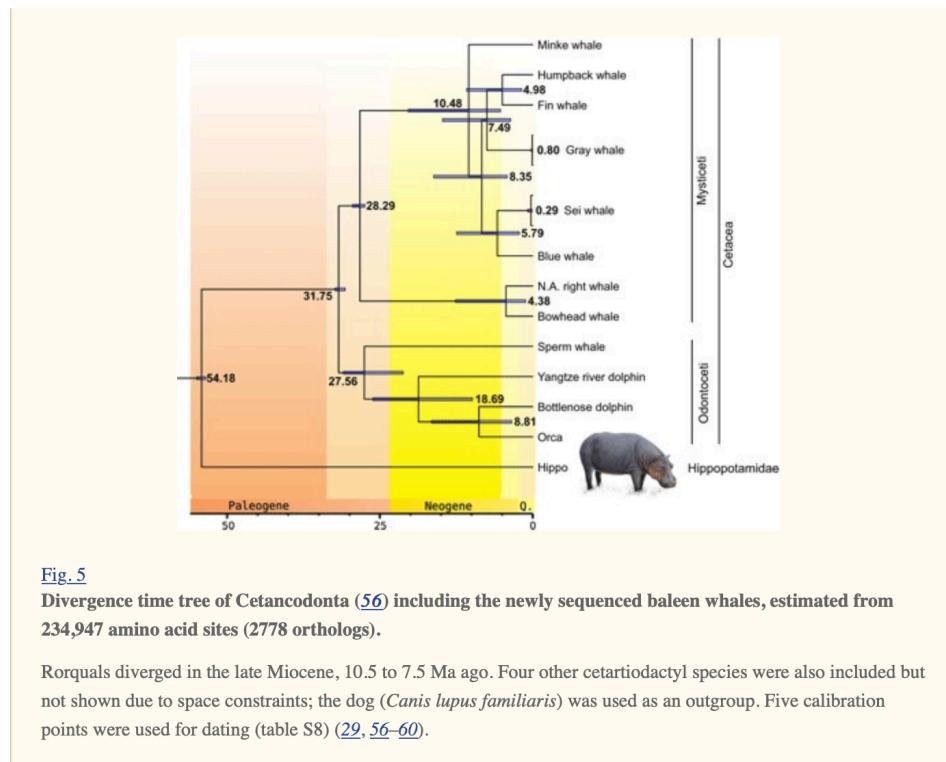
In sympatric speciation, genomes can be homogenized by gene flow, and only a few genes need to be under divergent selection to form new species .

Support hypotheses that selective processes maintain species divergence even under gene flow.

Science Advances | AAAS

SIGNIFICANT RESEARCH, GLOBAL IMPACT
Sci Adv. 2016 Apr; 4(4): ead9873.
Published online 2016 Apr 4.
doi: 10.1126/sciadv.069873

PMCID: PMC5084691
PMID: 26532802
Whole-genome sequencing of the blue whale and other rorquals finds signatures for introgressive gene flow
Olaf Amundsen,^{1,*} Esteri Lemmola,^{2,3,4,*} Vivas Kumar,² Maria A. Nilsson,² and Axel Janke^{2,3,4,†}



Canids show introgression in both presence and absence of sympatric speciation reflecting ancestral events during early divergence

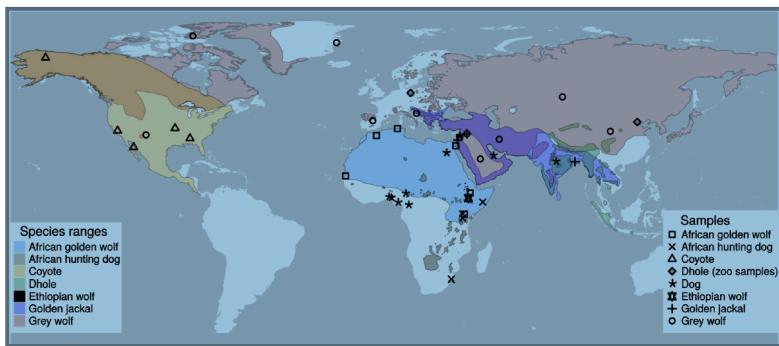
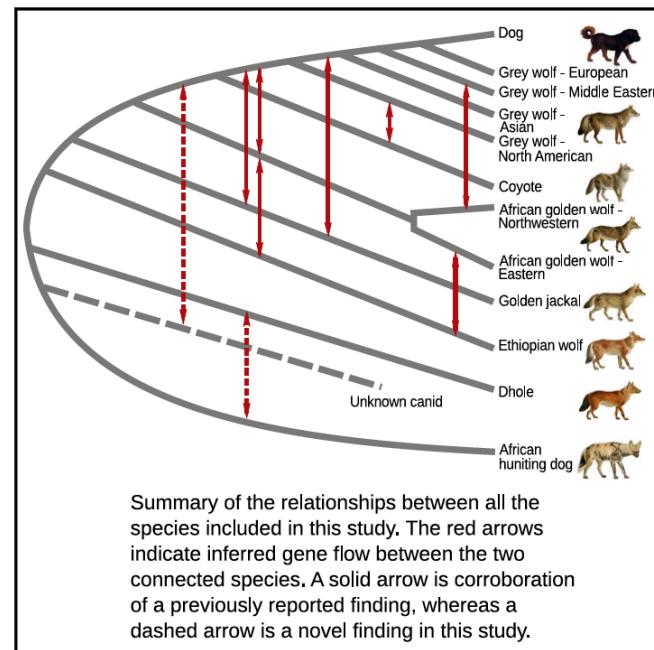


Figure 1. Map Showing the IUCN Ranges, Range Overlaps, and Sampling Locations of the Canids Included in This Study
The overlaps in ranges are shown in blended colors (orange, dark purple, dark olive green, light teal, etc.). Since IUCN does not have range information for African golden wolf, the IUCN range of golden jackal has been split in two; the Eurasian part is shown as the range of golden jackal, and the African part is shown as the range of African golden wolf. Further details on the samples, including their sampling location and source, can be found in [Data S1](#), and their estimated heterozygosities—which are inversely proportional to their population sizes—are shown in [Figure S1](#).

Interspecific Gene Flow Shaped the Evolution of the Genus *Canis*

Shyam Gopalakrishnan,^{1,21,22} Mikel-Holger S. Sinding,^{1,3,4,21} Jazmin Ramos-Madrigal,^{1,21} Jonas Niemann,¹ Jose A. Samaniego Castruita,¹ Filipe G. Vieira,¹ Christian Carpe,¹ Marc de Manuel Montero,⁵ Lukas Kuderna,⁶ Aitor Serres,⁷ Victor Manuel González-Basallo,⁸ Carlos Fernández,⁹ Philippe Gaubert,¹⁰ Klaus-Peter Koepfli,^{10,15} Jane Budd,¹⁰ Eli Knispel Ruiness,¹⁷ Mads Peter Heide-Jørgensen,¹² Bent Pedersen,^{13,14} Thomas Sicheritz-Ponten,^{16,19} Lutz Bachmann,¹⁸ Oystein Wiig,² Anders J. Hansen,^{1,24} and M. Thomas P. Gilbert^{1,20}



Gene tree heterogeneity in host genome: Opportunities for targeted genomics

Prior to advances in NGS and computational methods, phylogenomic analysis of 1000's of genes was inconceivable.

Now we can target genomic regions of interest based on their pattern of evolution within a species phylogeny.

Once the molecular phylogeny of the species of interest is resolved, the phylogeny serves as a constraint tree, or framework for interpreting evolution of targeted regions.

What is the target?

In wildlife disease dynamics, we are targeting the subset of gene trees linked with host survival during disease outbreaks.

coding regions = exome,

transcribed genes=transcriptome,

candidate gene lists from other sources

What are the basic steps for reaching our objectives in targeted genomics using phylogenomic methods?

A comprehensive sampling of the genetic diversity present in host taxa is essential as incomplete genomes, or missing taxa will undermine or bias the results

Class: Mammalia

Scenario:

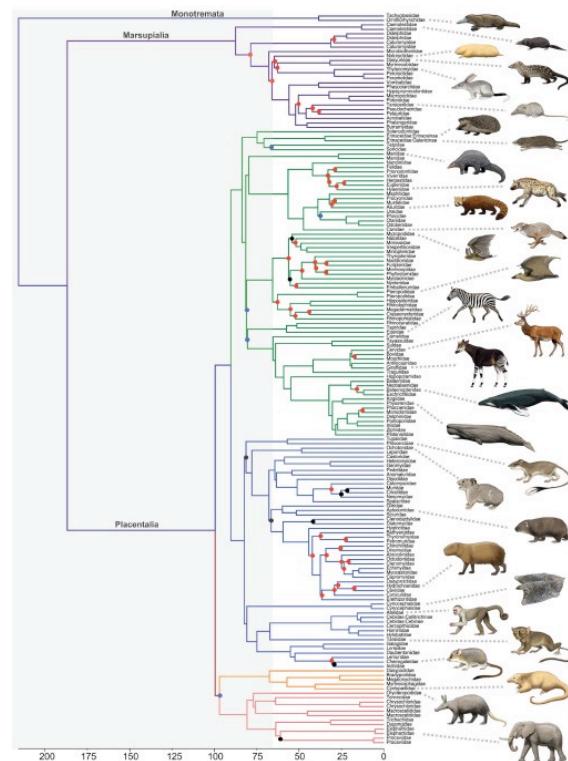
Pathogen infects a broad range of species from diverse lineages across Mammals

Phylogenomic Species Tree:

All mammalian Orders

NGS strategy design:

Targeted Genomics from representative species from all Order lineages



Order: Primates

Scenario:

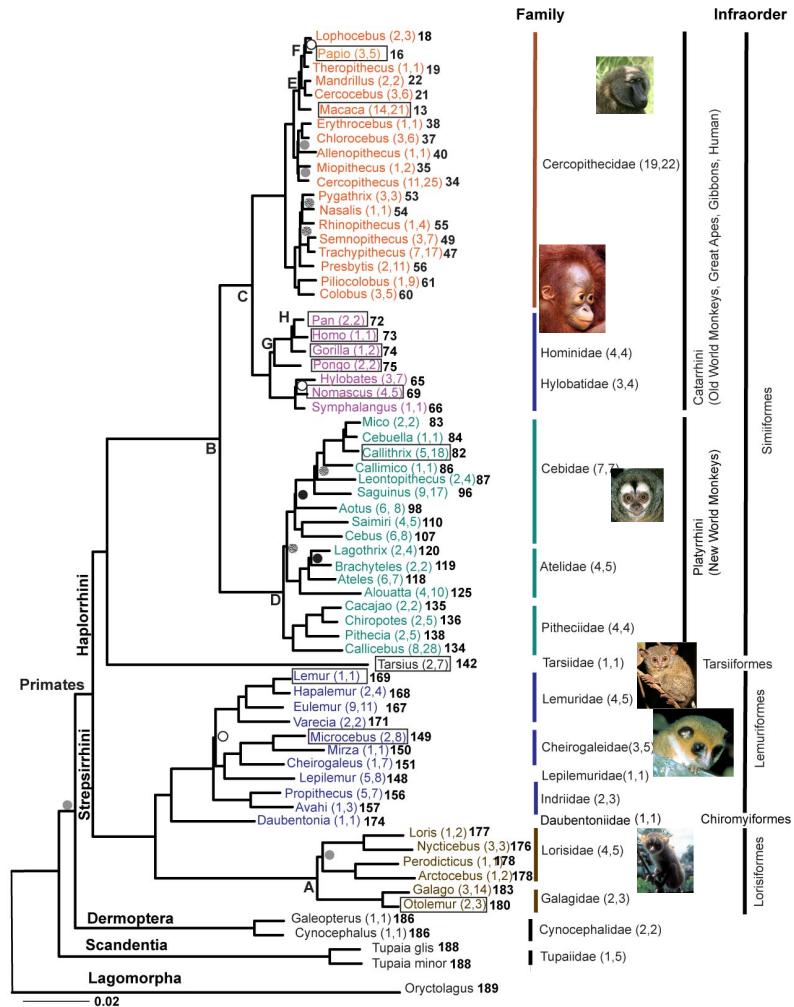
Pathogen infects species from diverse genera within Order

Phylogenomic Species Tree:

Genera (78) within Order

NGS Strategy and Design:

Targeted regions of representative species from each genus

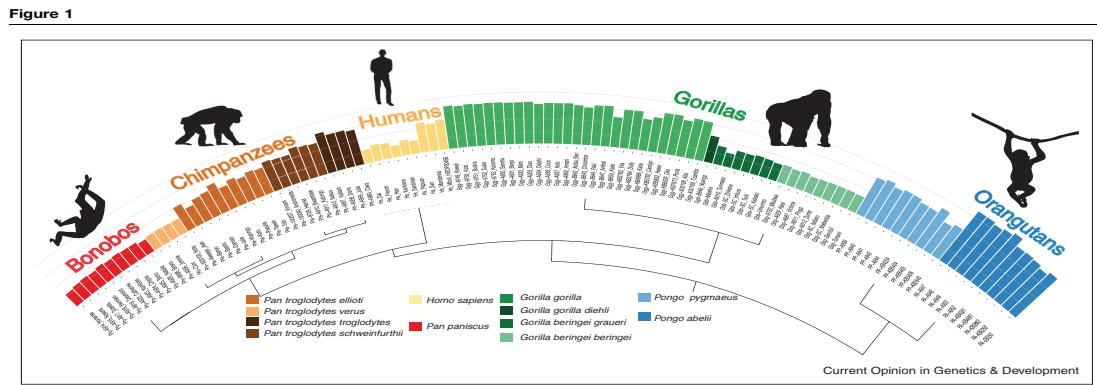


Species and Subspecies: Great Apes (Hominidae)

Scenario:
Pathogen infects specific subspecies

Phylogenomic Species Tree:
All species and subspecies within
Hominidae

NGS strategy design:
Targeted Genomics from multiple
individuals within species and
subspecies



Phylogenetic relationships and variation of genetic diversity among the great apes. The height of the bars represents genome-wide diversity. Source: Data from Locke *et al.* (2012), Prado-Martinez *et al.* (2013), and Xue *et al.* (2015) were re-analysed based on sequence read mapping to their respective species reference genomes.

Kuhlwilm, M. et al Evolution and demography of the great apes. Current Opinion in Genetics & Development 2016, 41:124–129

Species Phylogeography: Tarsiers

Scenario:

Pathogen infects specific populations of species range.

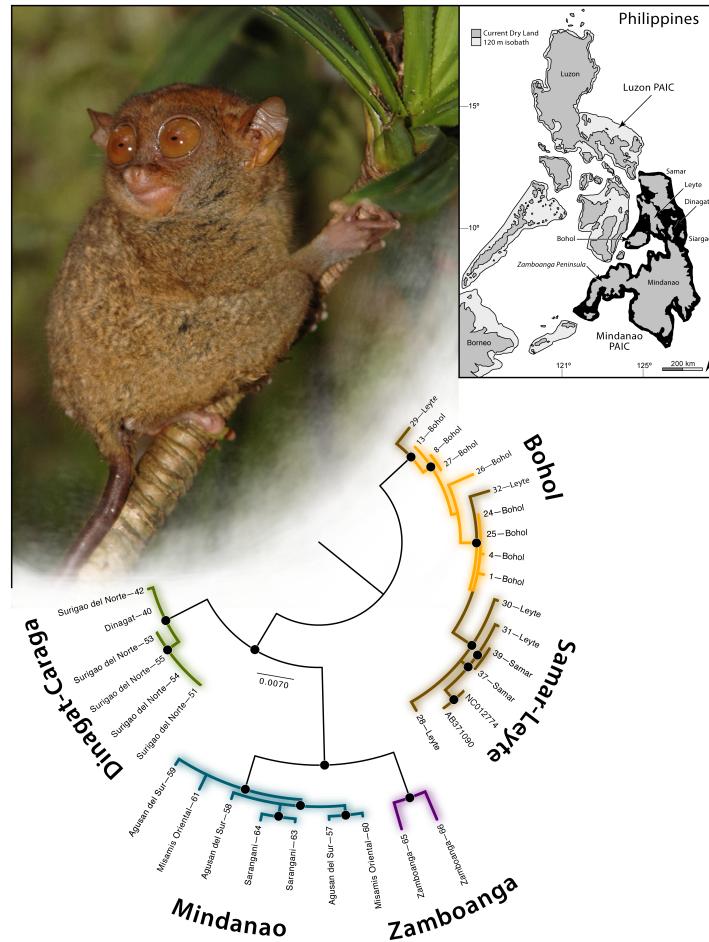
Phylogenomic Tree:

Evolution of populations across species range.

NGS strategy and design:

Targeted genomics of individuals within populations

Brown RM, Weghorst JA, Olson KV, Duya MRM, Barley AJ, et al. (2014) Conservation Genetics of the Philippine Tarsier: Cryptic Genetic Variation Restructures Conservation Priorities for an Island Archipelago Primate. PLOS ONE 9(8): e104340. <https://doi.org/10.1371/journal.pone.0104340>
<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0104340>



Phylogenomic Inferences of host targeted genomics and disease phenotype

- Reconstruct phylogenomic tree
- Assess total levels of gene tree heterogeneity
- Compare patterns of gene tree heterogeneity between resistant and susceptible lineages within phylogeny
- Some might define ongoing or past outbreaks of disease and episodic or adaptive evolution of host resistance/susceptibility
- Create candidate list of genes for detecting signature of selection in DNA variation relative to disease resistance/susceptibility